These graphs show average rewards over lifetime for different values of the exploration constant C.

C=0.01 has the runs with highest average rewards but also has many outliers runs where the run does not explore for actions with higher rewards but instead keeps exploiting their current optimal action.

C=0.1 has the runs with high average rewards without outliers. As C increases the runs become more varied with their average rewards.

C=1 has the runs with the lowest average rewards over lifetime because it explores too much and does not exploit what it has found to be the optimal actions.